



ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

Lexical Link Analysis (LLA) Application: Improving Web Service to Defense Acquisition Visibility Environment

30 September 2015

Dr. Ying Zhao, Research Associate Professor

Dr. Douglas J. MacKinnon, Research Associate Professor

Dr. Shelley P. Gallup, Research Associate Professor

Graduate School of Operational & Information Sciences

Naval Postgraduate School

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Abstract

We have been studying DoD acquisition decision making since 2009. The US DoD acquisition process is extremely complex. There are three key processes that must work in concert to deliver capabilities: determining warfighters' requirements and needs; planning the DoD budget, and procuring final products. Each process produces large amounts of information (Big Data). There is a critical need for automation, validation, and discovery to help acquisition professionals, decision makers and researchers understand the important content within large data sets and optimize DoD resources throughout the processes. Lexical Link Analysis (LLA) can help by applying automation to reveal and depict—to decision-makers—the correlations, associations, and program gaps across all, or subsets of, acquisition programs examined over many years. This enables strategic understanding of data gaps and potential trends, and can inform managers where areas might be exposed to higher program risk, and how resource and big data management might affect the desired return on investment (ROI) among projects. In this report, we describe new developments in analytics and visualization, how LLA is adaptive to Big Data Architecture and Analytics (BDAA), and reveal needs for Big Acquisition Data used in Defense Acquisition Visibility Environment (DAVE).

Keywords: Lexical Link Analysis, big data, big acquisition data, big data architecture, big data analytics



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Acknowledgments

We thank Mr. Robert Flowe from OUSD(AT&L)/ARA/OS&FM, who provided sponsorship to access the Defense Acquisition Visibility Environment(DAVE) along with relevant questions with many insightful discussions.



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

About The Authors

Dr. Ying Zhao is a Research Associate Professor at the Naval Postgraduate School and frequent contributor to DoD forums on knowledge management and data sciences. Her research and numerous professional papers are focused on knowledge management approaches such as data/text mining, Lexical Link Analysis, system self-awareness, Collaborative Learning Agents, search and visualization for decision-making, and collaboration. Dr. Zhao was principal investigator (PI) for six contracts awarded by the DoD Small Business Innovation Research (SBIR) Program. Dr. Zhao is a co-author of four U.S. patents in knowledge pattern search from networked agents, data fusion and visualization for multiple anomaly detection systems. She received her Ph.D. in mathematics from MIT and is the Co-Founder of Quantum Intelligence, Inc.

Dr. Ying Zhao
Information Sciences Department
Naval Postgraduate School
Monterey, CA 93943-5000
Tel: 831-656-3789
Fax: (831) 656-3679
E-mail: yzhao@nps.edu

Dr. Doug MacKinnon is a Research Associate Professor at the Naval Postgraduate School (NPS). Dr. MacKinnon leads multi-disciplinary studies ranging from leading the Analyst Capability Working Group (ACWG) for the U.S. Air Force in 2014, studying Maritime Domain Awareness (MDA) in 2010, as well as Knowledge Management (KM) and Lexical Link Analysis (LLA) projects. He also led the assessment for the Tasking, Planning, Exploitation, and Dissemination (TPED) process during the Empire Challenge 2008 and 2009 (EC08/09) field experiments and for numerous other field experiments of new technologies during Trident Warrior 2012 (TW12). He teaches courses in operations research (OR) and holds a PhD from Stanford University, conducting successful theoretic and field research in Knowledge Management (KM). He has served as the program manager for two major government projects of over \$50 million each, implementing new technologies while reducing manpower requirements. He has served over 20 years as a naval surface warfare officer, amassing over eight years at sea and serving in four U.S. Navy warships with five major, underway deployments.

Dr. Douglas J. MacKinnon
Information Sciences Department and Graduate School of Operational and Information Sciences
Naval Postgraduate School
Monterey, CA 93943-5000
Tel: 831-656-1005
Fax: (831) 656-3679
E-mail: djmackin@nps.edu



Dr. Shelley Gallup is a Research Associate Professor at the Naval Postgraduate School's Department of Information Sciences, where he directs major field experimentation studies leveraging numerous at-sea Navy exercises. . Dr. Gallup has a multidisciplinary science, engineering, and analysis background, including microbiology, biochemistry, space systems, international relations, strategy and policy, and systems analysis. He returned to academia after retiring from naval service in 1994 and received his PhD in engineering management from Old Dominion University in 1998. Dr. Gallup joined NPS in 1999, bringing his background in systems analysis, naval operations, military systems, and experimental methods first to the Fleet Battle Experiment series (1999–2002) and then to Fleet experimentation in the Trident Warrior series (2003–2014). Dr. Gallup's interests are in Knowledge Management (KM) and complex systems field experimentation.

Dr. Shelley P. Gallup
Information Sciences Department
Naval Postgraduate School
Monterey, CA 93943-5000
Tel: 831-656-1040
Fax: (831) 656-3679
E-mail: spgallup@nps.edu





ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

Lexical Link Analysis (LLA) Application: Improving Web Service to Defense Acquisition Visibility Environment

30 September 2015

Dr. Ying Zhao, Research Associate Professor

Dr. Douglas J. MacKinnon, Research Associate Professor

Dr. Shelley P. Gallup, Research Associate Professor

Graduate School of Operational & Information Sciences

Naval Postgraduate School

Disclaimer: The views represented in this report are those of the author and do not reflect the official policy position of the Navy, the Department of Defense, or the federal government.



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Table of Contents

Executive Summary	xiii
Background.....	1
Methodology	3
Lexical Link Analysis (LLA)	3
Word Pairs Generalization and CCC Method.....	6
References	19



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

- x -

List of Figures

Figure 1. DoD Acquisition Decision Making	xiii
Figure 2. Themes Discovered in Colored Groups	3
Figure 3. A Detailed View of a Theme in Figure 2	4
Figure 4. An Example of Emerging Theme	5
Figure 5. An Example of Anomalous Theme.....	5
Figure 6. Diagram for the LLA Method	6
Figure 7. The Word Pairs/Bi-gram in LLA a Context-Concept-Cluster (CCC) Model.....	6
Figure 8. Diagram for the CCC Method	7
Figure 9. Defense Acquisition Visibility Environment (DAVE)	9
Figure 10. Example LLA Reports and Visualizations	10
Figure 11. D3 Visualization Bar and Pie Charts	10
Figure 12. D3 Visualization: Collapsible Tree	11
Figure 13. D3 Visualization: Word Cloud.....	12
Figure 14. D3 Visualization: Hierarchical Edge Bundling	13



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Executive Summary

We have been studying DoD acquisition decision making (Gallup, MacKinnon, Zhao, Robey & Odel, 2009; Zhao, Gallup & MacKinnon, 2010, 2011, 2012a, 2012b, 2013 & 2014) since 2009. The US DoD acquisition process is extremely complex. There are three key processes that must work in concert to deliver capabilities: definition of warfighters' requirements and needs; DoD budget planning, and procurement of products, as in Figure 1. Each process produces increasingly large volumes of information (Big Data). The need for automation, validation, and discovery is now a critical need, as acquisition professionals, decision makers and researchers grapple to understand data and make decisions to optimize DoD resources.

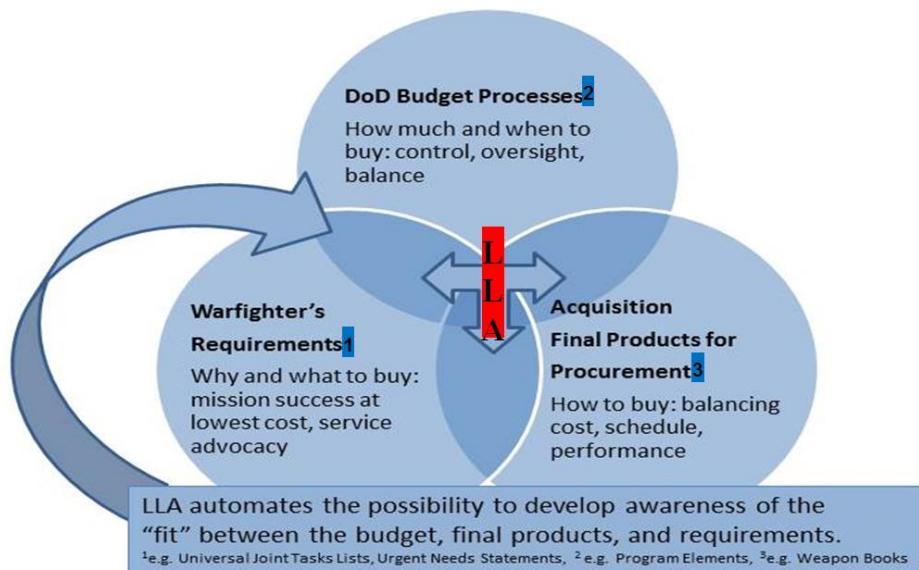


Figure 1.DoD Acquisition Decision Making

We have been working on the problem of how the interlocking systems processes become aware of their fit between DoD programs and warfighters' needs and how can gaps of non-fit be revealed? Moreover, in the performance of DoD acquisition processes, each functional community is required to review only the particular information for which they are responsible, further exacerbating the problem of lack of fitness. For example, the systems engineering community typically only examines the engineering documents and feasibility studies, the test and evaluation community looks only at the test and evaluation plans, and the acquisition community looks only at acquisition strategies. Rarely do these stakeholders review each other's data or jointly discuss the core questions and integrated processes together as depicted in Figure 1.



In this report, we describe new developments in analytics and visualization, how LLA is adaptive to Big Data Architecture and Analytics (BDAA), and reveal needs for Big Acquisition Data used in Defense Acquisition Visibility Environment (DAVE). This enables: 1) discovery topics and themes; 2) comparison, sorting and ranking data the importance of the data source; 3) strategic understanding of the correlations of associations among data attributes, data vocabularies and data entities; 4) discovery of potential trends, patterns, anomalies and data gaps; and 5) correlation of the findings with areas that might have higher program risk, and how resources might be managed for the desired Return On Investment (ROI).



Background

Motivated by the lack of fit and horizontal integration in the DoD acquisition process, we have been applying Lexical Link Analysis (LLA), a data-driven automation technology and methodology across DoD acquisition processes to:

- Surface themes and their relationships across multiple data sources.
- Discover high value areas for investment.
- Compare and correlate data from multiple data sources.
- Sort and rank important and interesting information.

LLA is a data-driven method for pattern recognition, anomaly detection and data fusion. It shares indexes not data, and is therefore feasible for parallel and distributed processing, adaptive to Big Data Architecture and Analytics (BDAA) and capable of analyzing Big Acquisition Data.

As a motivating example from past work, we conducted a detailed examination of the Research, Development, Test and Evaluation (RDT&E) budget modification practice from one year to the next, over the course of ten years and 450 DoD Program Elements. We found a pattern revealing that the programs with fewer links (measured by LLA) to warfighters' requirements, received more budget reduction in total but less on average, indicating that budget reduction may have focused mainly on large and expensive programs rather than perhaps cutting all the programs that do not match warfighters' requirements. Furthermore, the programs with more links to each other received more budget reduction in total, as well as on average, indicating a pattern of good practice of allocating DoD acquisition resources to avoid overlapping efforts and to fund new and unique projects. These findings were useful as validation and guidance for future decision processes for automatically identifying programs to match warfighter's requirements, limit overall spending, maximize efficiencies, eliminate unnecessary costs and maximize the return on investment.



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Methodology

Lexical Link Analysis (LLA)

LLA has been used to analyze unstructured and structured data for pattern recognition, anomaly detection, and data fusion. It uses the theory of System Self-Awareness (SSA) to identify high-value information in the data that can be used to guide future decision processes in a data-driven or unsupervised learning fashion. It is implemented via a smart infrastructure named “system and method for knowledge pattern search from networked agents (US patent 8,903,756)” also known as Collaborative Learning Agents (CLA), licensed from Quantum Intelligence, Inc. (Zhou, Zhao, and Kotak, 2009).

In LLA, a complex system is expressed in specific vocabularies or lexicons to characterize its features, attributes or its surrounding environment. LLA uses bi-gram word pairs as the features to form word networks. Figure 2 depicts using LLA to analyze ten years of the reports in the NPS Acquisition Program with word pairs as groups or *themes*. Figure 3 shows a detail of a *theme* in Figure 2. A node represents a word. A link or edge represents a word pair.

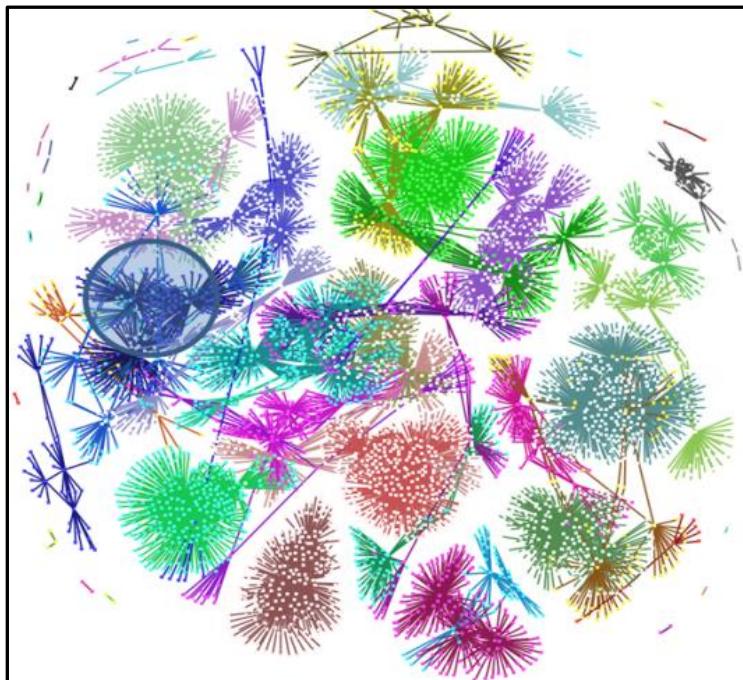


Figure 2. Themes Discovered in Colored Groups



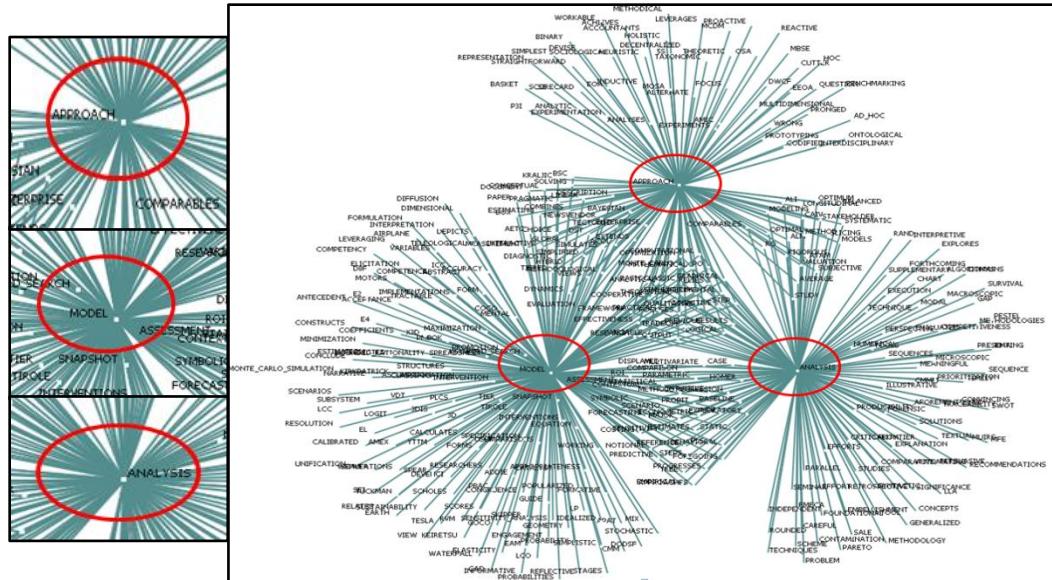


Figure 3. A Detailed View of a Theme in Figure 2

LLA is related to bags-of-words (BAG) methods such as LDA (Blei, Ng & Jordan, 2003) and text-as-network (TAN) methods such as the Stanford Lexical Parser (SLP; SNLPG, 2015). LLA selects and groups features into three basic types:

- Popular (P): Main themes found from the data. Figure 3 is an example of a popular theme centered around word nodes “analysis, model, and approach.” These themes could be less interesting because they are already in the public consensus and awareness. They represent the patterns in the data.
 - Emerging (E): Themes that may grow to be popular over time. Figure 4 is an example of an emerging theme centered around word nodes “national, defense, and acquisition.”
 - Anomalous (A): Themes that may be off-topic but that may be interesting for further investigation. Figure 5 is an example of anomalous theme centered around word nodes “stock, and market(s).”



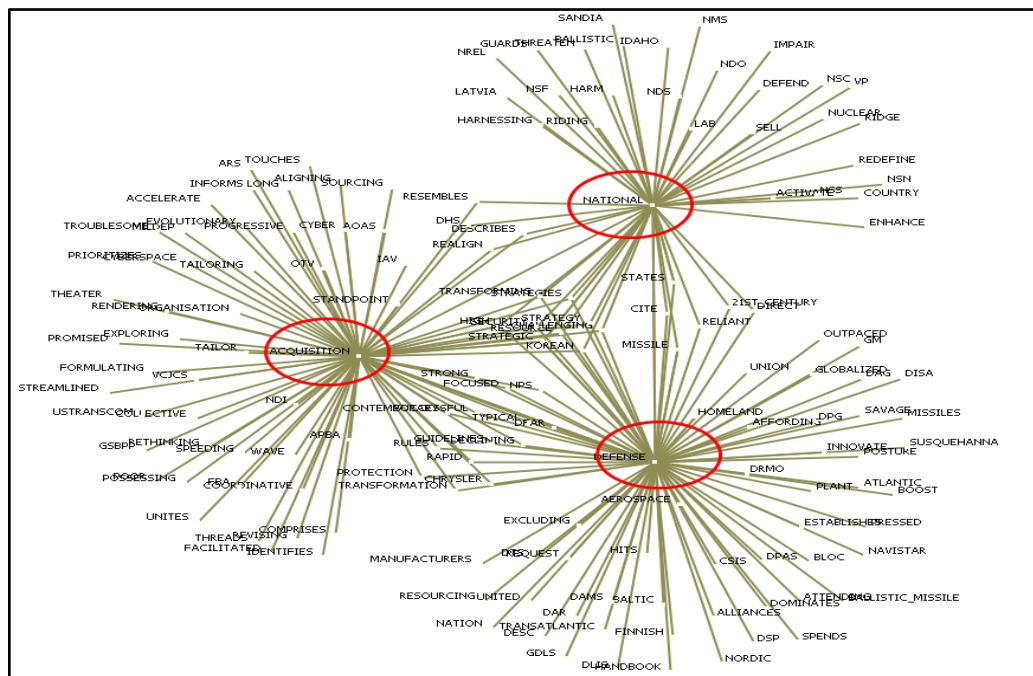


Figure 4. An Example of Emerging Theme

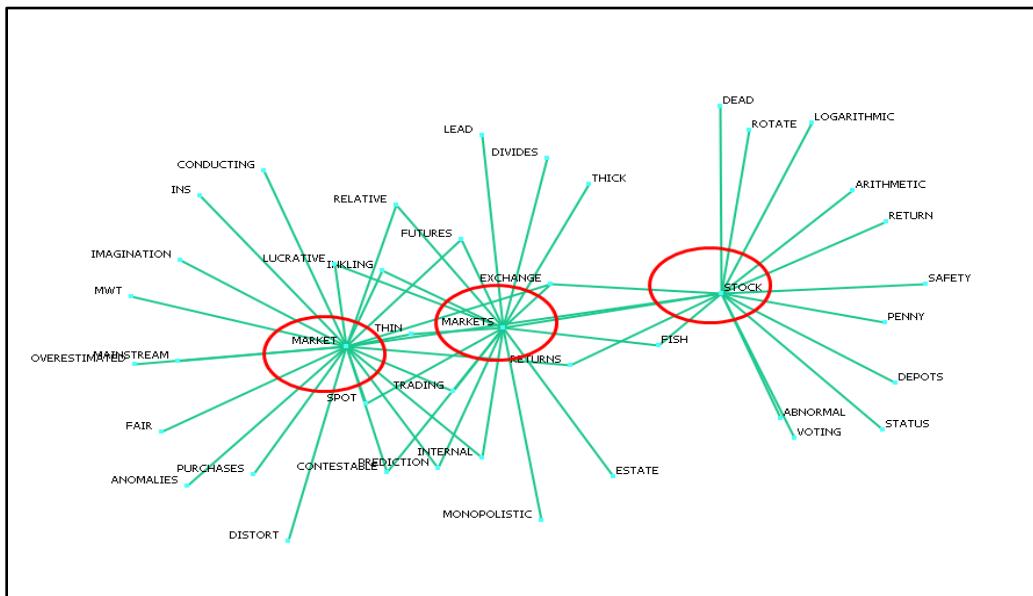


Figure 5. An Example of Anomalous Theme

Figure 6 summarizes LLA used for historical and new data. The red part shows a pattern (e.g., a theme) discovery phase using historical data including data fusion that come from multiple learning agents. The black part shows an application phase that a new data is compared with the patterns discovered and hence the anomalies are revealed.



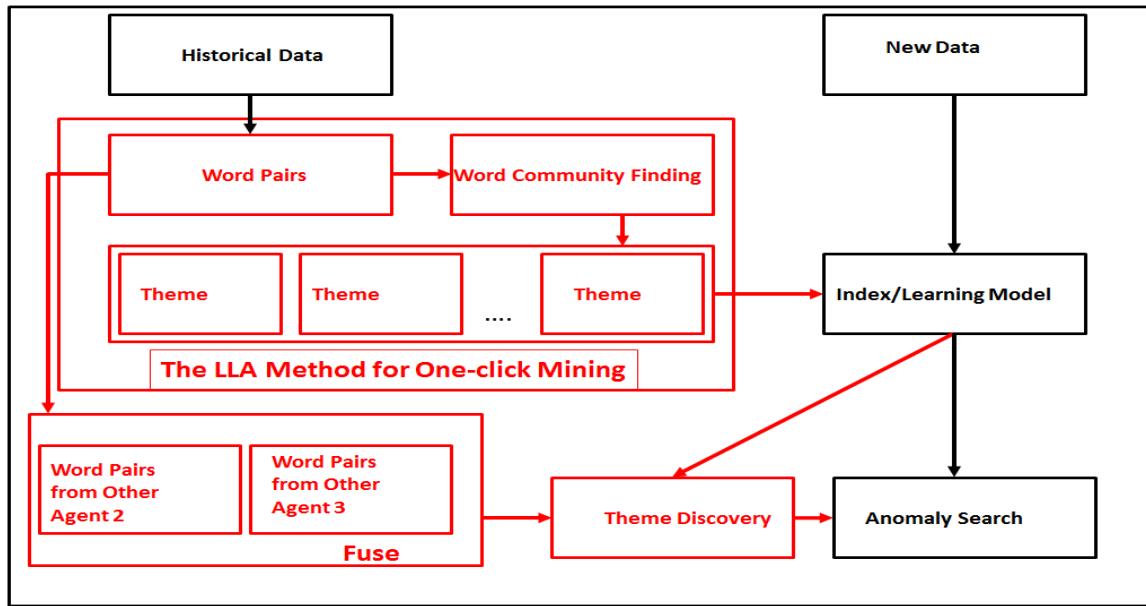


Figure 6. Diagram for the LLA Method

Word Pairs Generalization and CCC Method

Figure 7 shows the word pairs/bi-gram in an LLA can be generalized as a Context-Concept-Cluster (CCC) model, where a context is a generic attribute which can be shared by multiple data sources, a concept is a specific attribute for a data source and a cluster is a combination of attributes or themes that can be computed using a word community finding algorithm (e.g., Girvan & Newman, 2002) in Figure 6 to characterize a data set. Context can be a word, location, time or object, etc.

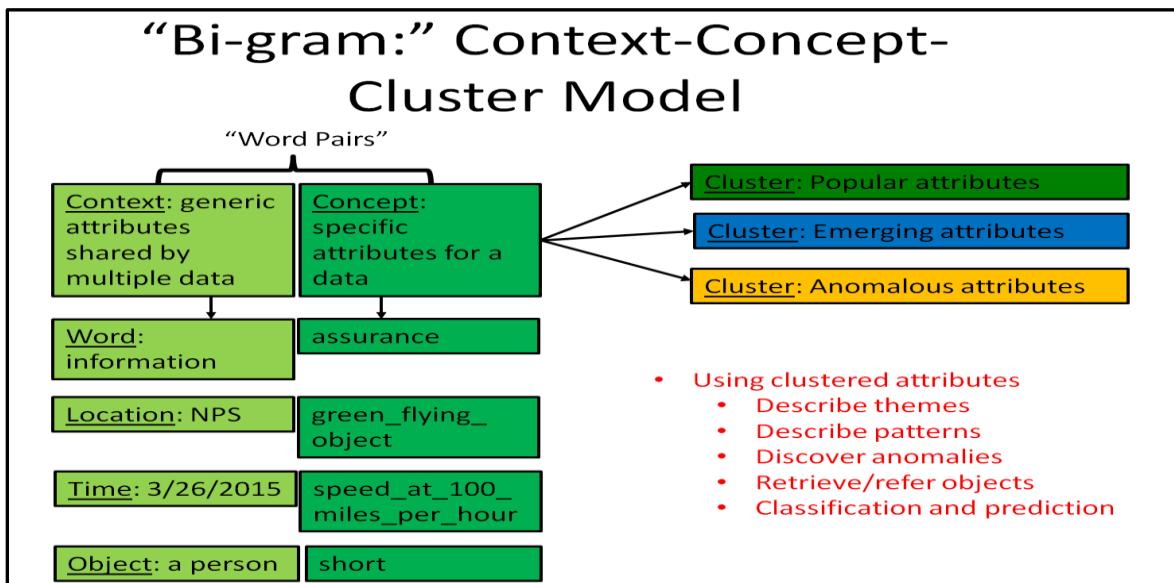


Figure 7. The Word Pairs/Bi-gram in LLA a Context-Concept-Cluster (CCC) Model



Figure 8 summarizes how a generalized CCC method is used for historical and new data. Similar to Figure 6 there is a pattern discovery phase using historical data where patterns are learned and discovered, and an application phase for a new data is compared with the patterns discovered and anomalies are revealed.

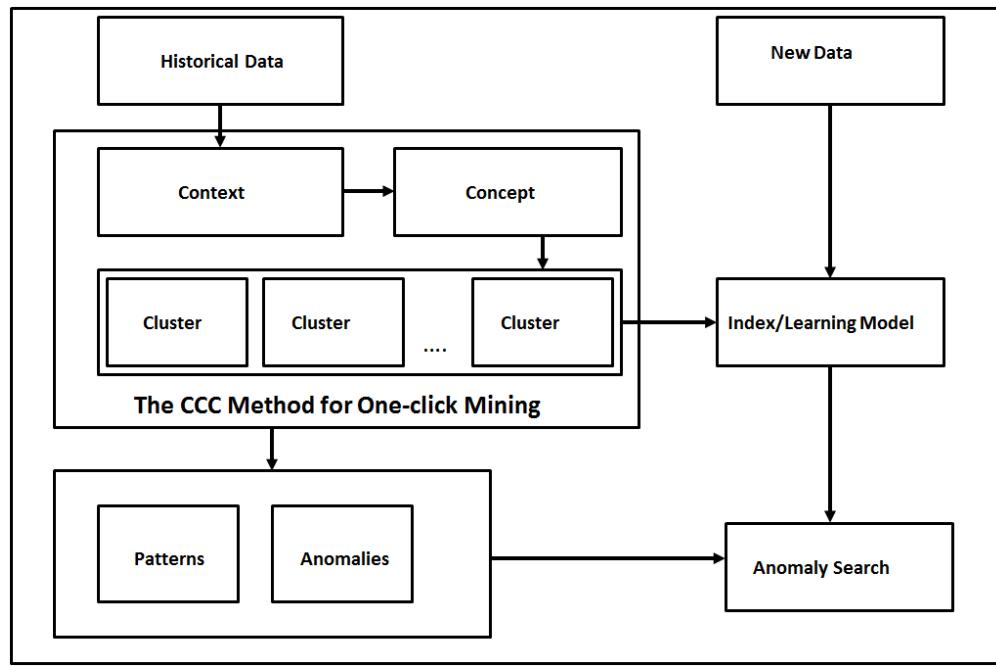


Figure 8: Diagram for the CCC Method



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Research Results

In this section, we describe the research results for the following two tasks completed this year.

Task 1. We worked with the OSD OUSD ATL (US) to install the LLA/SSA/CLA system as a web service in the Defense Acquisition Visibility Environment (DAVE) test bed. It was approved, installed and in progress for source code security scanning. The goal has been to bring the tool close to the DAVE sources as shown in Figure 9.

- DAVE interfaces with authoritative data sources to support the OUSD(AT&L) and his staff in decision-making, oversight and analysis for major programs
- Sample data sources
 - **Defense Technical Information Center (DTIC)**
 - Program Elements (PEs):
[\[http://www.dtic.mil/descriptivesum/\]](http://www.dtic.mil/descriptivesum/)
 - **Defense Acquisition Management Information Retrieval (DAMIR)**
(<http://www.acq.osd.mil/damir/>)
 - Selected Acquisition Report (SAR)
 - Defense Acquisition Executive Summary (DAES)
 - Acquisition Program Baseline (APB)
 - **Acquisition Information Repository (AIR)**
 - Test & Evaluation Master Plan (TEMP)
 - Systems Engineering Plan (SEP)
 - Life Cycle Sustainment Plan (LCSP)
 - Acquisition Strategy Reports (ASR)
 - Acquisition Decision Memorandum (ADM, for Milestone B 2366b Certification)
 - Technology Readiness Assessment (TRA), and LCSP

Figure 9. Defense Acquisition Visibility Environment (DAVE)

Figure 10 shows the existing LLA reports and visualizations in the LLA installed. The details of the existing visualizations can be found in the earlier publications (Zhao, Gallup & MacKinnon, 2013, 2014 & 2015). Figure 11, 12, 13, and 14 are the new visualizations developed this year using the Data-Driven Documents (D3) tool. Figure 11 shows multiple bar and pie charts for multiple themes discovered using LLA when comparing two data sources. Specifically, it reveals: how many lexical terms are matched (red) and how many lexical terms are unique for each data source (green) and in each theme. Figure 12 shows the lexical terms in each theme in a collapsible tree. Figure 13 shows the lexical terms in a word cloud with each color representing a theme. Figure 14 shows clusters of data sources in a



hierarchical edge bundling. The clusters are generated based on the numbers of matched and unique lexical terms found among the data sources.

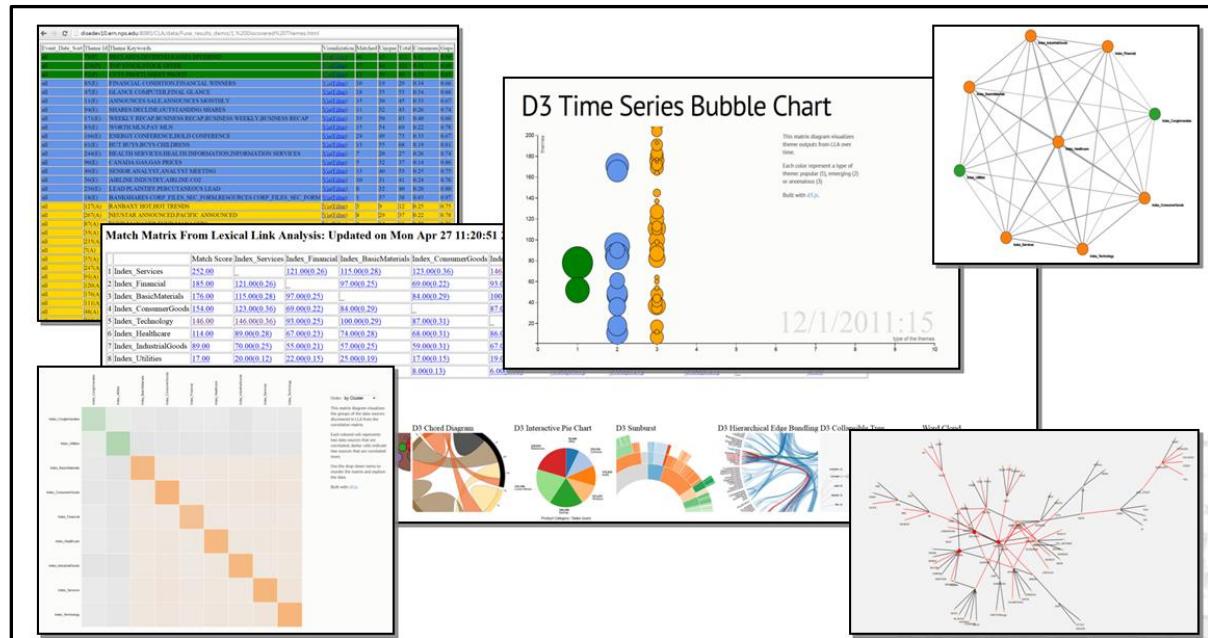


Figure 10. Example LLA Reports and Visualizations

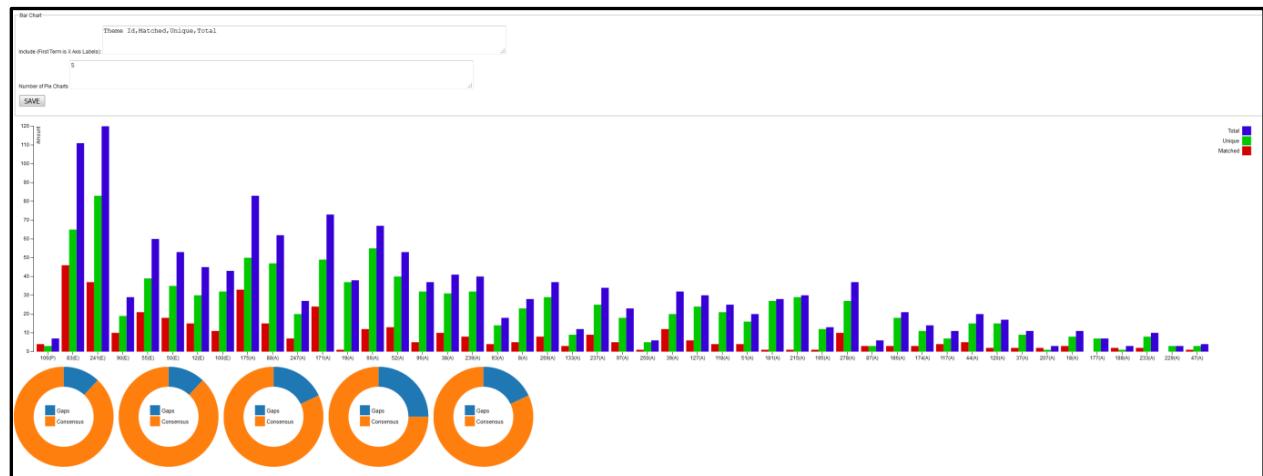


Figure 11. D3 Visualization Bar and Pie Charts: The Number Lexical Terms Matched in Red and Number of Lexical Terms Unique in Green



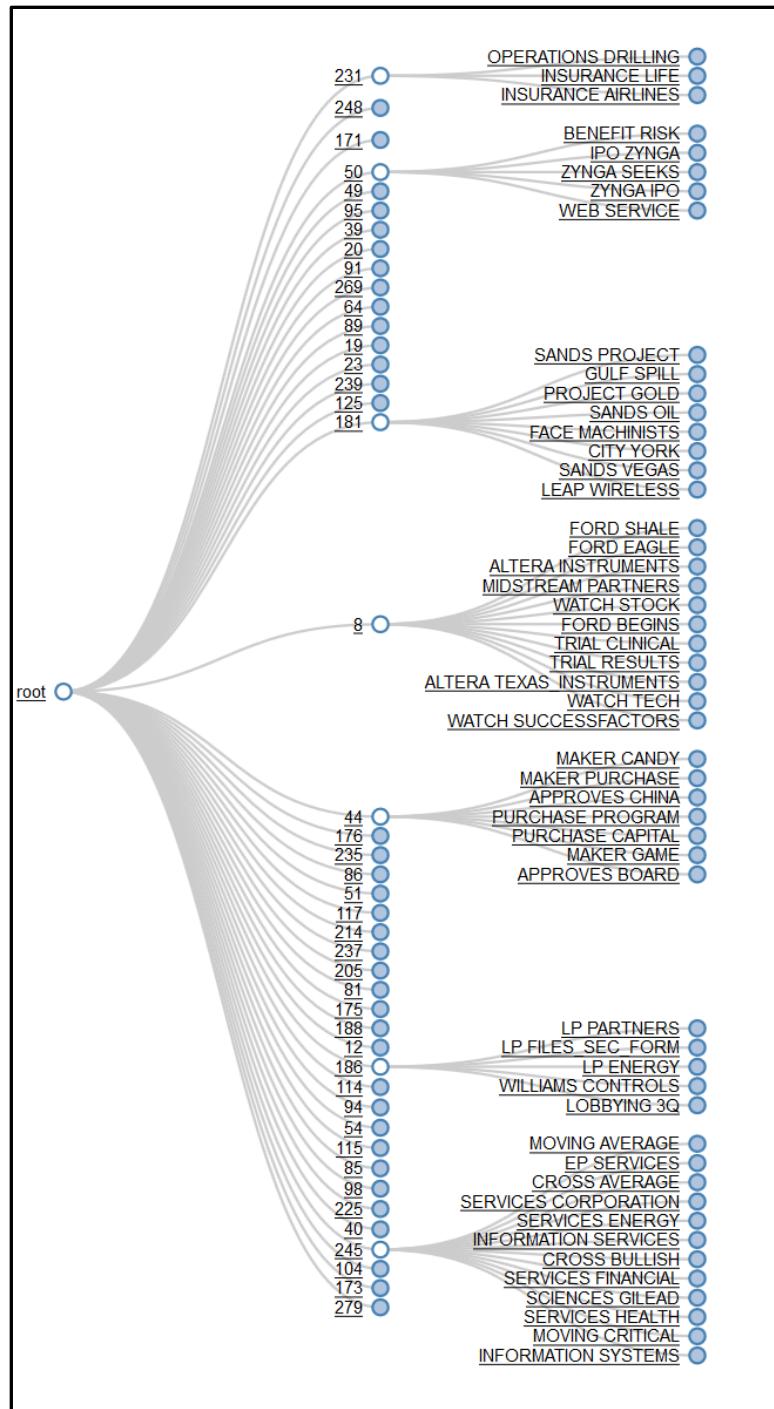


Figure 12. D3 Visualization - Collapsible Tree: Lexical Terms in Each Theme Listed in an Interactive and Collapsible Tree



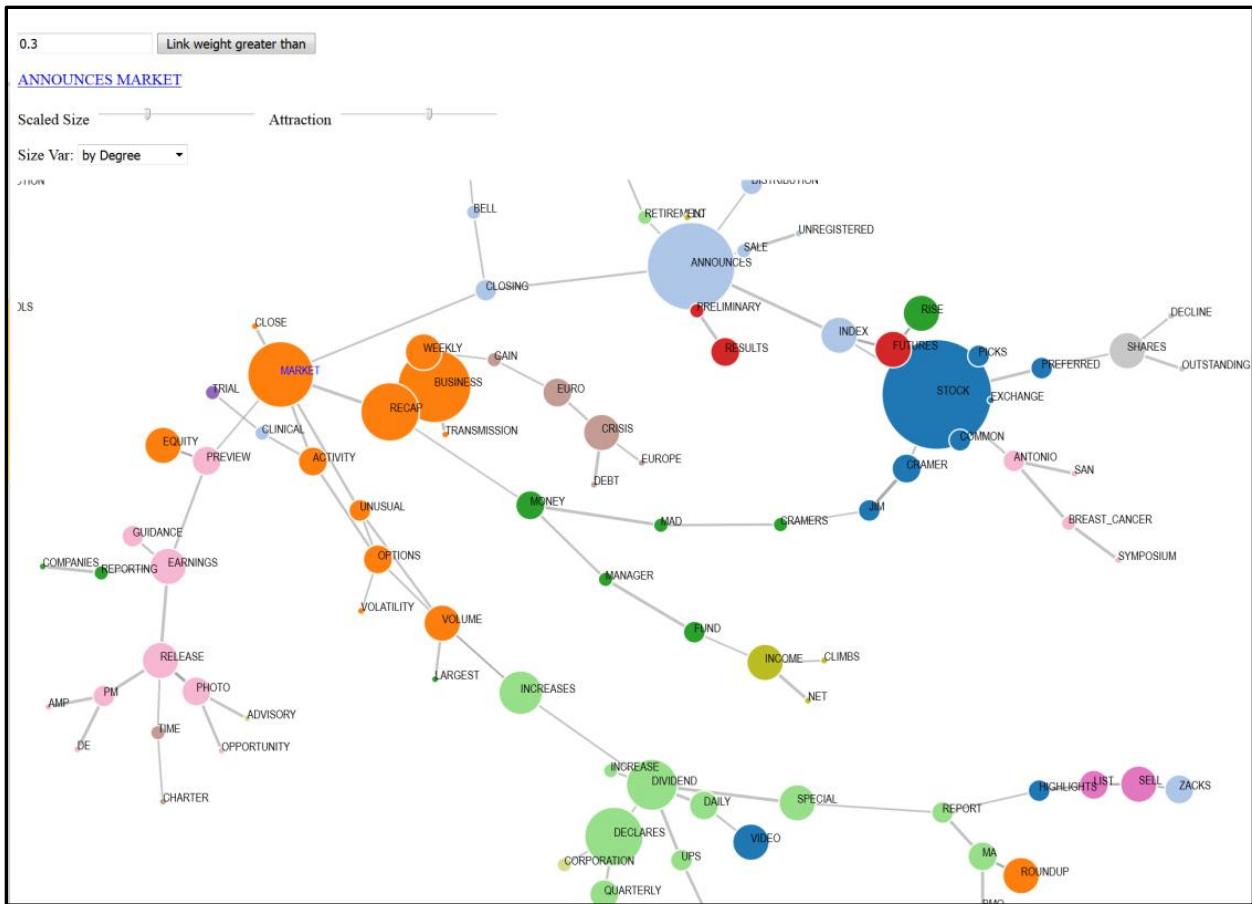


Figure 13. D3 Visualization - Word Cloud: Each Color Representing a Theme, Filters and Interactions on Link Weight, Scale and Attraction



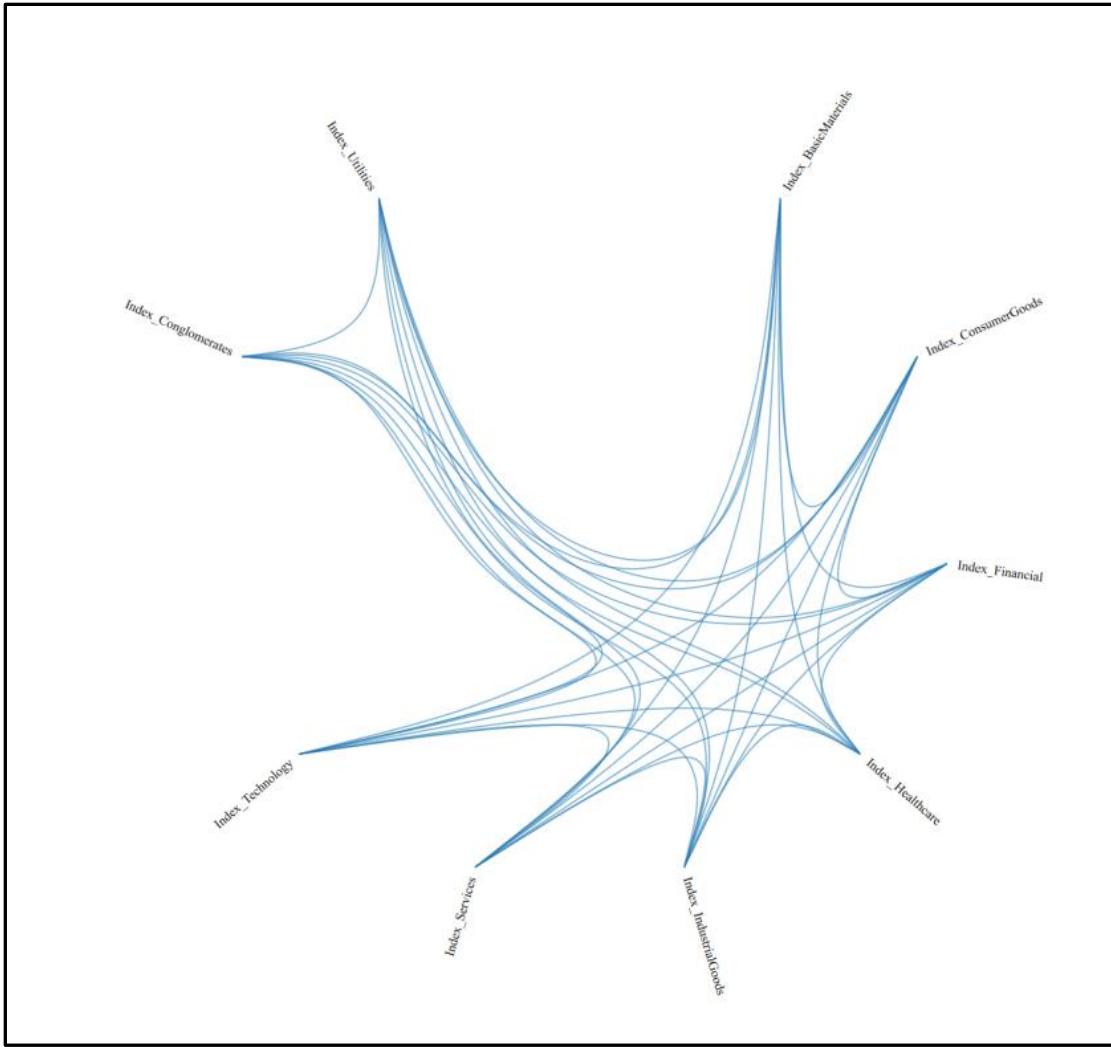


Figure 14. D3 Visualization - Hierarchical Edge Bundling: Showing Clusters of Data Sources (e.g., Index_Technology, etc.). The Clusters and Links Among Data Sources Measured by the Number of Lexical Terms

Task 2. We also explored how to use LLA jointly with other business intelligence tools especially Big Data Architecture and Analytics (BDAA) tools:

Transfer learning, unsupervised learning and deep learning across heterogeneous data sources are trends for the academic and commercial Big Data applications. Deep Learning includes unsupervised machine learning techniques (e.g., neural networks) for recognizing objects of interest from Big Data, for instance, sparse coding (Olshausen & Field, 1996) and self-taught learning (Raina, Battle, Lee, Packer,& Ng,2007). The self-taught learning approximates the input for unlabeled objects as a succinct, higher-level feature representation of sparse linear combination of the bases. It uses the Expectation and Maximization (EM) method to iteratively learn coefficients and bases. Deep Learning links machine vision and text analysis smartly. For example, text analysis Latent Dirichlet Analysis (LDA) is a



sparse coding where a bag of words used as the sparsely coded features for text (Olshausen & Field, 1996). Our methods Lexical Link Analysis (LLA), System-Self-Awareness (SSA), and Collaborative Learning Agents (CLA) can be viewed as unsupervised learning or Deep Learning for pattern recognition, anomaly detection, and data fusion.

We explored a few deep learning techniques; in particular, a recursive data fusion methodology leveraging LLA, SSA, and CLA was examined as follows:

- An agent j represents a sensor, operates on its own like a decentralized data fusion, however it does not communicate with all other sensors but only with the ones that are its peers. A peer list can be specified by the agent.
- An agent j includes a learning engine CLA that collects, analyzes from its domain specific data knowledge base $b(t,j)$, for examples, $b(t,j)$ may represent the statistics for bi-gram feature pairs (word pairs) computed from LLA.
- An agent j also includes a fusion engine SSA with two algorithms SSA1 and SSA2 that can be customized externally. SSA1 integrates the local knowledge base $b(t,j)$ to the total knowledge base $B(t,j)$ that can be passed along to its peers and used globally in the recursion in Figure 6. SSA2 assesses the total value of the agent j by separating the total knowledge base into the categories of patterns, emerging and anomalous themes based on the total knowledge base $B(t,j)$ and generates a total value $V(t,j)$ as follows:
 - Step 1: $B(t,j) = \text{SSA1}(B(t-1, p(j)), b(t,j))$;
 - Step 2: $V(t,j) = \text{SSA2}(B(t,j))$

Where $p(j)$ represents the peer list of agent j .

- The total value $V(t,j)$ is used in the global sorting and ranking of relevant information.

In this recursive data fusion, the knowledge bases and total values are completely data-driven and automatically discovered from the data. Each agent has the exact same code of LLA, SSA, and CLA, yet has its own data apart from other agents. This agent work has the advantages of both decentralized and distributed data fusion. It performs learning and fusion simultaneously and in parallel. Meanwhile, it categorizes the patterns and anomalous information.

We have also met the acquisition professionals and discussed how LLA and BDAA can be applied to the DoD acquisition process. The following list is a summary of business applications showing how Big Data analytic tools such as LLA can help decision maker's understanding of Big Acquisition Data:

- Examine and compare the important acquisition data sources to gain business insights to:
 - Study prime and subcontractor relationships.



- Analyze DOT&E Annual Reports using horizontal analysis.
- Compare DOT&E and DAES: What did the OT results reveal and did issues surface?
 - Did these issues show up somewhere in DAES earlier?
- Perform Time series analysis.
 - Can LLA look at program ratings over time?
- Compare budget data and contracts data.
- Other areas of Big Data exploration may include:
 - Understanding how in the current acquisition process, a small delay or anomaly in a contract negotiation process can have a huge impact in its performance, therefore cost the government a lot of money downstream.
 - Applying BDAA such as LLA for pattern recognition and anomaly detection for these kind of problems to enable early warnings and predictions to prevent the downstream risks.
 - Leveraging Big Acquisition Data to include programs' cost, SAR, DAMIR, technical data, to perhaps include outside economic environment data if access is possible.
 - Learning causes of the deviations from the normal behaviors for the programs/contracts and how we might use models arising from physics (e.g., fluid dynamics theories).
 - Exploring LLA's network perspectives and social dynamics found among the nodes and the System Self-Awareness (SSA) theory and how they may be used to lay out the academic rigor for business processes, for example, to answer the following questions:
 - Are some nodes drawn towards some other nodes because the other nodes are more powerful, for instance, the powerful nodes can represent the organizations with more decision makers?
 - Should the growth pattern be attached to popular existing ideas (i.e., preferential attachment), or innovative ideas (i.e., expertise growth)? How are the forces of the nodes modeled and mapped into the social network settings and actual business processes?



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Conclusions

On September 17, 2015, we met with Mr. Mark E. Krzysko, Deputy Director, Acquisition, Resources and Analysis Enterprise Information at the OUSD (AT&L). In that discussion, we agreed that future work will also include installing LLA at the Defense Manpower Defense Center (DMDC); a *Federally Funded Research and Development Center [FFRDC]*) and setting up a BDAA consortium, whereby DAVE could provide an authoritative yet secure data throughput to the DMDC and NPS students, faculty, and acquisition research professionals. This in turn, can inspire a large community of researchers and NPS students, who may be future leaders of the DoD acquisition community, to use the Big Data tools and analytics like LLA to access the DAVE data to answer important research questions and advance the state-of-the-art of DoD acquisition.

As a proof-of-concept, we installed the LLA tool in the OUSD(AT&L) e-business center to access multiple categories of program data as authentic and authoritative data sources in DAVE. We also enhanced the LLA analytical capabilities, reports and visualizations to show it is a tool for Big Data Acquisition for understanding the insight of the current acquisition business processes, answer important business questions that lead to the identification of future specific and productive directions for acquisition research.

We are at the onset of many exciting advances in understanding more about the DoD acquisition system by leveraging Big Data methodology including LLA and we look forward to further in-depth engagement in the years to come.



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

References

- Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. Retrieved from <http://jmlr.csail.mit.edu/papers/volume3/blei03a/blei03a.pdf>
- Gallup, S. P., MacKinnon, D. J., Zhao, Y., Robey, J., & Odell, C. (2009, October 6–8). Facilitating decision making, re-use and collaboration: A knowledge management approach for system self-awareness. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management (IC3K)*, Madeira, Portugal.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, USA*, 99(12), 7821–7826.
- Olshausen, B. & Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*.
- Quantum Intelligence (QI). (2009). *Collaborative learning agents (CLA)*. Retrieved from <http://www.quantumii.com/qi/cla.html>
- Raina, R., Battle, A., Lee, H., Packer, B., & Ng, A.Y. (2007). Self-taught learning: Transfer learning from unlabeled data. In ICML.
- Stanford Natural Language Processing Group (SNLPG) (2015). Retrieved from <http://nlp.stanford.edu/software/lex-parser.shtml>, Spark (2015). Retrieved from <http://spark.apache.org/>
- Zhou, C., Zhao, Y., & Kotak, C. (2009). The Collaborative Learning agent (CLA) in Trident Warrior 08 exercise. In KDIR, Madeira, Portugal, INSTICC Press.
- Zhao, Y., Gallup, S.P., & MacKinnon, D.J. (2010). Towards real-time program awareness via lexical link analysis. *Acquisition Research Sponsored Report Series, NPS-AM-10-174*. Retrieved from <http://www.acquisitionresearch.net/files/FY2010/NPS-AM-174.pdf>
- Zhao, Y., MacKinnon, D.J., Gallup, S.P. (2011). Lexical Link Analysis for the Haiti earthquake relief operation using open data sources. In *Proceedings of the 6th ICCRTS, International Command and Control, Research and Technology Symposium*, Québec City, Canada June 21–23, 2011. Retrieved from http://www.dodccrp.org/events/16th_iccrts_2011/papers/164.pdf
- Zhao, Y., Gallup, S.P., & MacKinnon, D.J. (2011). System self-awareness and related methods for improving the use and understanding of data within DoD. *Software Quality Professional*, 13(4), 19-31. Retrieved from



<http://www.nps.edu/Academics/Schools/GSOIS/Departments/IS/DISE/docs/improving-use-and-understanding-of-data-dod.pdf>

Zhao, Y., Gallup, S.P., Mackinnon, D.J. (2011). A web service implementation for large-scale automation, visualization, and real-time program-awareness via lexical link analysis. *Acquisition Research Sponsored Report Series, NPS-AM-11-186*. Retrieved from <http://www.acquisitionresearch.net/files/FY2011/NPS-AM-11-186.pdf>

Zhao, Y., MacKinnon, D., & Gallup, S. (2012a). Semantic and social networks comparison for the Haiti earthquake relief operations from APAN data sources using lexical link analysis. In *Proceedings of the 17th ICCRTS, International Command and Control, Research and Technology Symposium*, Fairfax, Virginia, June 19-21, 2012. Retrieved from http://www.dodccrp.org/events/17th_iccrts_2012/post_conference/papers/082.pdf

Zhao, Y., Gallup, S.P., Mackinnon, D.J. (2012b). Applications of lexical link analysis web service for large-scale automation, validation, discovery, visualization and real-time program-awareness. *Acquisition Research Sponsored Report Series, NPS-AM-12-205*. Retrieved from <http://www.acquisitionresearch.net/files/FY2012/NPS-AM-12-205.pdf>

Zhao, Y., Gallup, S., Mackinnon, D.J. (2013). Lexical Link Analysis application: Improving web service to acquisition visibility portal. *Acquisition Research Sponsored Report Series, NPS-AM-13-109*. Retrieved from <http://www.acquisitionresearch.net/files/FY2013/NPS-AM-13-109.pdf>

Zhao, Y., Gallup, S., Mackinnon, D.J. (2014). Lexical link analysis application: improving web service to acquisition visibility portal phase II. In *Proceedings of the eleventh annual acquisition research symposium*, Monterey, California, May, 2014. Retrieved from <http://www.acquisitionresearch.net/files/FY2014/NPS-AM-14-C11P17R03-065.pdf>

Zhao, Y., Gallup, S., Mackinnon, D.J. (2015). Lexical link analysis application: improving web service to acquisition visibility portal phase III. In *Proceedings of the eleventh annual acquisition research symposium*, Monterey, California, May, 2015. Retrieved from <http://www.acquisitionresearch.net/files/FY2015/SYM-AM-15-108.pdf>





ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net